View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Portal de Revistas de Nicaragua

Nexo Revista Científica
Vol. 33, No. 02, pp. 431-445/Diciembre 2020
ISSN-E 1995-9516
Universidad Nacional de Ingeniería
COPYRIGHT © (UNI). TODOS LOS DERECHOS RESERVADOS
http://revistas.uni.edu.ni/index.php/Nexo
https://doi.org/10.5377/nexo.v33i02.10782
CC BY NC ND

Forecast of the efficiency confidence interval of decision-making units in data envelopment analysis

Pronóstico del intervalo de confianza en la eficiencia de las unidades de toma de decisiones en el análisis envolvente de datos

Azarnoosh Kafi[1], Behrouz Daneshian[2,*], Mohsen Rostamy-Malkhalifeh[3]

[1] Department of Mathematics, Lahijan Branch, Islamic Azad University, Lahijan, Iran.

[2] Department of Mathematics, Central Tehran Branch, Islamic Azad University, Tehran, Iran.

[3] Department of Mathematics, Science and Research Branch, Islamic Azad University, Tehran, Iran.

*Corresponding author email: be_daneshian@yahoo.com

(recibido/received: 03-July-2020; aceptado/accepted: 01-September-2020)

ABSTRACT

Data Envelopment Analysis (DEA) is a well-known method for calculating the efficiency of Decision-Making Units (DMUs) based on their inputs and outputs. When the data is known and in the form of an interval in a given time period, this method can calculate the efficiency interval. Unfortunately, DEA is not capable of forecasting and estimating the efficiency confidence interval of the units in the future. This article, proposes a efficiency forecasting algorithm along with 95% confidence interval to generate interval data set for the next time period. What’s more, the manager’s opinion inserts and plays its role in the proposed forecasting model. Equipped with forecasted data set and with respect to data set from previous periods, the efficiency for the future period can be forecasted. This is done by proposing a proposed model and solving it by the confidence interval method. The proposed method is then implemented on the data of an automotive industry and, it is compared with the Monte Carlo simulation methods and the interval model. Using the results, it is shown that the proposed method works better to forecast the efficiency confidence interval. Finally, the efficiency and confidence interval of 95% is calculated for the upcoming period using the proposed model.

Keywords: Data Envelopment Analysis, Forecast, Time Series, Probability Programming, Efficiency, Monte Carlo Simulation, Confidence Interval.

RESUMEN

El análisis envolvente de datos (DEA) es un método bien conocido para calcular la eficiencia de las unidades de toma de decisiones (DMU) en función de sus entradas y salidas. Cuando los datos son conocidos y en forma de intervalo en un período de tiempo dado, este método puede calcular el intervalo de eficiencia. Desafortunadamente, la DEA no es capaz de pronosticar y estimar el intervalo de confianza de eficiencia de las unidades en el futuro. Este artículo propone un algoritmo de pronóstico de eficiencia junto con un intervalo de confianza del 95% para generar un conjunto de datos de intervalo para el próximo período de tiempo. Además, la opinión del gerente se inserta y desempeña su papel en el modelo de pronóstico propuesto. Equipado con un conjunto de datos pronosticado y con respecto al conjunto de datos de períodos anteriores, se puede pronosticar la eficiencia para el período futuro. Esto se hace proponiendo un modelo propuesto y resolviéndolo mediante el método del intervalo de confianza. A continuación, el método propuesto se implementa sobre los datos de una industria automotriz y se compara con los métodos de simulación de Monte Carlo y el modelo de intervalo. Usando los resultados, se muestra que el método propuesto funciona mejor para pronosticar el intervalo de confianza de eficiencia. Finalmente, se calcula la eficiencia y el intervalo de confianza del 95% para el próximo período utilizando el modelo propuesto.

Palabras clave: Análisis envolvente de datos, pronóstico, series de tiempo, programación de probabilidad, eficiencia, simulación de Montecarlo, intervalo de confianza.

1. INTRODUCCIÓN

Data Envelopment Analysis (DEA) is a non-parametric technique which has been used to determine the relative efficiency of the Decision-Making Units (DMUs) at a time when their inputs and outputs are known and congruent. CCR was first introduced by Charles et al. 1984 in order to determine the relative efficiency of units. After that, Banker et al. 1978 presented the BCC model, and later more generalizations were made on these two basic models. On the other hand, considering the uncertainty, we sometimes encounter data in the DEA, which is not precisely known, but their values are within a certain interval, such as the amount of budget, revenues, costs, etc. The IDEA model or interval DEA model accepts data in an interval and it is assumed that the data of each decision-making unit is within a given interval. Cooper et al. 1999,2001 first examined how to deal with unspecified data such as bounded data. Finally, Wang et al. 2005 is determined for each DMU with bounded interval, which is determined by the best efficiency of upper and lower bounds. Jahanshahloo et al. 2009 presented IGDEA as a general model with interval data for the interval DEA (IDEA), which can use IDEA's basic models with integrated interval data. In addition, they showed the theoretical properties of the relationships between the IGDEA and IDEA models.

The flaw of standard DEA models speaks out on previous data. In other words, DEA standard models fail in forecasting procedure and can evaluate efficiency with historical data. There are various papers which contribute forecasting procedure employing linear programming and DEA techniques. Fildes et al. 2011 used a three-stage combined forecast method to forecast the short-term energy efficiency of a region over the next six years. Xu et al. 2012 evaluated the relative performance of crude oil price forecast models based on data envelopment analysis. Lim et al. 2014 used the Technology Forecast Data Envelopment Analysis (TFDEA) method. Then Lim et al. 2015 used the TFDEA to forecast the technology of supercomputers development and measure their changes. Emrouznejad et al. 2016 used DEA models to rank several forecasting techniques and they calculated the error of the methods for the comparison of the accuracy of different forecast methods. For further discussion, among the researches refer to Zerafat Angiz 2012, Hatami-marbini 2011, Barak et.al 1984, Shabanpour et.al 2017, Peykani et.al 2019, and Tavassoli et.al 2019.

Obviously, the management of each decision-making unit is interested in forecasting the efficiency interval in the upcoming period, so that it regulates its activities in terms of resource consumption and output, but there are situations in which the data is not certainly within the given range. The main hint in these works is ignoring the decision maker's opinion. These works provoked a question in real world application, which was wondered in this paper. What would happened if the role of managers are proposed in forecasting procedure? As far as we know, in real world competitions, managers have to ponder various factors such as dynamics of environmental factors, resource consumption and output production. Besides, they have to rely on information about the period when the units under review have passed this period. Admittedly, the obtained results based on the past data cannot lead to such desirable outcomes. Although, generalization of the results do not allow the managers to adjust the production activities in terms of resources and productions. Hence, the managers interferes their viewpoints into evaluation which can affect the final results seriously. With reference to historical data along with decision maker's features, the efficiency of units can be forecasted. Another point which was argued in this paper is confidence. In order to have trustable results, this paper claims confidence interval of efficiency. A 95% confidence interval for input/output measures are resulted with an alternative algorithm. Equipped with manager's arguments and confidence interval of data sets, the efficiency is forecasted.

In this paper, a method is proposed to forecast an efficiency interval in the foregoing periods considering the known efficiency and inputs and outputs of each decision-making unit in previous periods. Because DEA models are not able to forecast efficiency in upcoming periods with known data records of decision making units. The main argument of this paper is that the data with certain probability is in the given intervals, for which we propose a model and solve it with a confidence interval method. We then use the proposed method to forecast the efficiency in an automobile industry and compare it with Monte Carlo simulation and interval model. Finally, the efficiency value and efficiency confidence interval of 95% are calculated for the upcoming period using the proposed model. At the end, a real case of 20 automobile industries with utilizing the last 49 periods can support the suggested approach in this paper.

The remainder of this paper is organized as follows:

Section 2 briefly introduces classic data envelopment analysis, Interval Data envelopment analysis and Monte Carlo method. Section 3 describes the The proposed method. A real case of numerical example is demonstrated in Section4. Section 5 concludes the paper.

2. PRELIMINARIES

2.1. Efficiency Analysis

In the classic data envelopment analysis, the following pattern is used to represent inputs and outputs of the decision-making unit.

Figure 1. The classic pattern of the decision-making unit

x_{io} and y_{ro} are positive integers, respectively, which are representing the i-th input and r-th output of DMU_o. Assume x_{ij} and y_{rj} , $(i = 1,2,\dots,m\ ;\ j = 1,2,\dots,n\ ;\ r = 1,2,\dots,s)$ are the input and output values of n decision-making unit. The envelopment form of CCR model is used to determine the efficiency as follows:

$$\begin{aligned} Min \quad & \theta_o \\ s.t \quad & \sum_{j=1}^{n} \lambda_j x_{ij} \le \theta_o x_{io} \\ & \sum_{j=1}^{n} \lambda_j y_{rj} \ge y_{ro} \qquad (1) \\ & \lambda_j \ge 0 \qquad j = 1,\dots,n \end{aligned}$$

2.2. Interval Data envelopment analysis (IDEA)

Considering the uncertainty, DEA sometimes encounters unknown data, including when a set of DMUs includes interval data, ordinal data, unknown data, or fuzzy data. In many applied cases, data is in an interval. Therefore, model (1) is not appropriate. In this case, the following interval pattern is used.

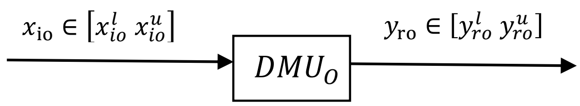

Figure 2. Interval pattern

Assume that the inputs and outputs of DMU_j are within an interval where, x_{ij}^l and x_{ij}^u u are respectively the lower and upper bounds of the i-th input and, y_{rj}^l and y_{rj}^u are respectively the lower and upper bounds of the r-th output of DMU_j. As there are $x_{ij}^l \leq x_{ij} \leq x_{ij}^u$ and $y_{rj}^l \leq y_{rj} \leq y_{rj}^u$ and also, $x_{ij}^l \leq x_{ij}^u$ and $y_{rj}^l \leq y_{rj}^u$. Such data is called interval data.

In this way, the lower and upper efficiency bounds of the evaluated unit of DMU_o are obtained using the following two models.

$$
\begin{aligned}
& Max \quad \theta_o^l = \sum_{r=1}^{s} u_r y_{ro}^l \\
& s.t. \quad \sum_{i=1}^{m} v_i x_{io}^u = 1 \\
& \sum_{r=1}^{s} u_r y_{ro}^l - \sum_{i=1}^{m} v_i x_{io}^u \leq 0 \\
& \sum_{r=1}^{s} u_r y_{rj}^u - \sum_{i=1}^{m} v_i x_{ij}^l \leq 0 \quad j = 1,2,\dots,n \quad , \quad j \neq o \\
& v_i \geq 0 \qquad i = 1,2,\dots,m \\
& u_r \geq 0 \qquad r = 1,2,\dots,s
\end{aligned} \tag{2}
$$

Where, the unit under evaluation are in the worst state and other units are in the best state for the calculation of θ_o^{l*}.

$$
\begin{aligned}
& Max \quad \theta_o^u = \sum_{r=1}^{s} u_r y_{ro}^u \\
& s.t. \quad \sum_{i=1}^{m} v_i x_{io}^l = 1 \\
& \sum_{r=1}^{s} u_r y_{ro}^u - \sum_{i=1}^{m} v_i x_{io}^l \leq 0 \\
& \sum_{r=1}^{s} u_r y_{rj}^l - \sum_{i=1}^{m} v_i x_{ij}^u \leq 0 \quad j = 1,2,\dots,n \quad , \quad j \neq o \\
& v_i \geq 0 \qquad i = 1,2,\dots,m \\
& u_r \geq 0 \qquad r = 1,2,\dots,s
\end{aligned} \tag{3}
$$

In this case, in order to calculateθ_o^{u*}, the unit under evaluation is in the best position and other units are in their worst position. Then, by computing θ_o^{l*} and θ_o^{u*}, we will have an interval as $[\theta_o^{l*}, \theta_o^{u*}]$which provides all possible efficiency measures for the unit under evaluation.

2.3 Monte Carlo method

In this study, one of the methods for solving the proposed model for comparison is using Monte Carlo method. Suppose n decision making units are available in accordance with pattern 3. The following flowchart is presented to determine the inputs and outputs of DMU_o ($o \in \{1, \dots, n\}$) and compute its efficiency by performing Monte Carlo simulations:

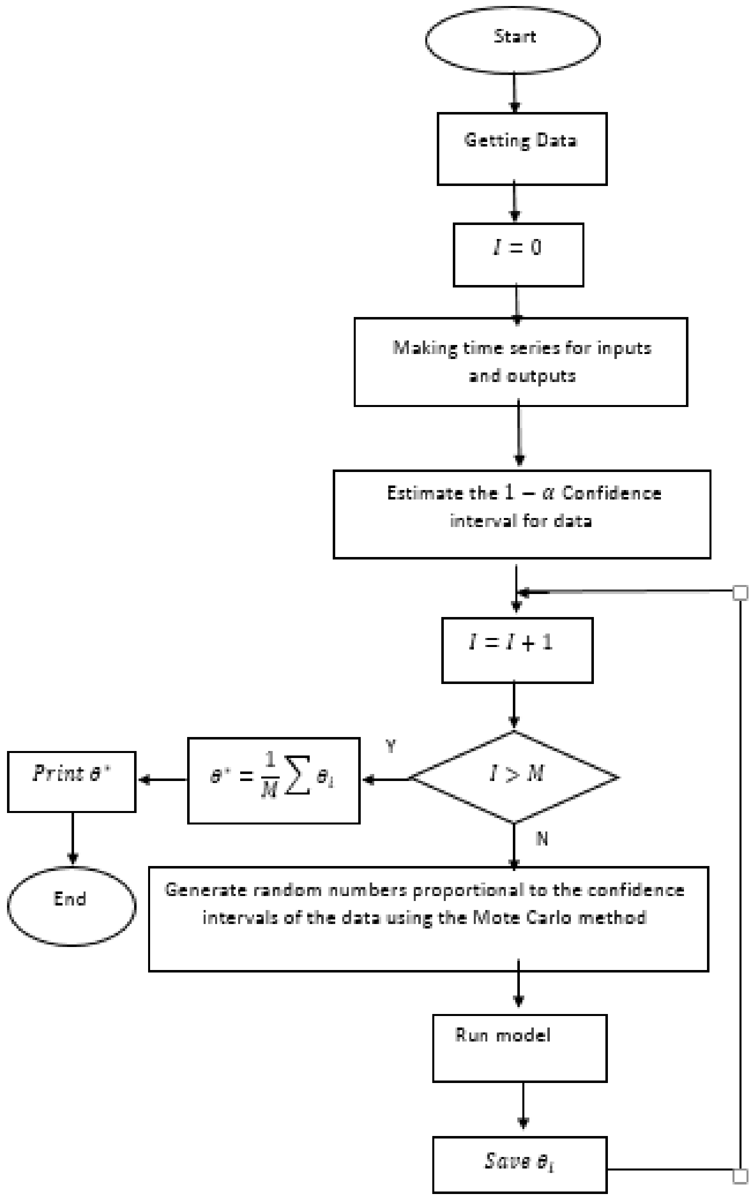

Figure 3. Flowchart of Monte Carlo Solution Method

3. PROPOSED APPROACH

Concerning the increasing trend of models for estimating efficiency, we understand that accurate value of data (inputs and outputs) was a known and non-negative number in the early version, i.e. inputs and outputs of a unit are expressed accurately and correctly by a number. Moreover, sometimes in the real world, we are faced with decision-making units that model in figure (1) cannot be used for them. In DEA, we are sometimes faced with unknown data concerning lack of absoluteness, for example, when a set of DMUs includes interval data.

For example, assume a factory in which budget is an input for it. It is evident that the budget cannot be a number due to operational or job issues or unpredicted events; thus, the budget is given in form of an interval or the number of personnel of the factory is in an interval and is not fixed concerning the amount of demand or season. The proposed model in figure (4) which is a generalization of the interval model of figure (2), is presented as follows.

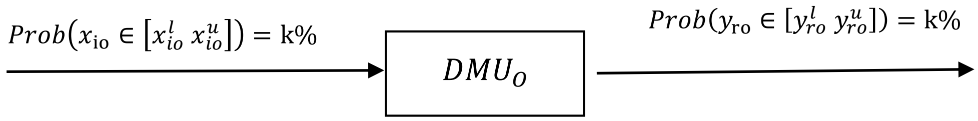

Figure 4. Proposed Pattern

As the figure (4) demonstrates, the $i-th$ component of input vector for DMU_o is meeting the interval $[x_{io}^{L}, x_{io}^{U}]$ with the possibility of $k\%$. In a similar manner, $r-th$ component of output vector is inserted in interval $[y_{ro}^{L}, y_{ro}^{U}]$ with the $k\%$ possibility. That is to say, with the possibility of $(1-k)\%$ the data is not registered in the interval.
In most cases, the confidence level, k is equal to 95%. Based on the pattern in Figure (4), the following model is applied to forecasting the interval efficiency while the confidence level 95%. The model has the following format:

$$
\begin{aligned}
&\text{Min } \theta_o \\
&S.t. \quad \sum_{j=1}^{n} \lambda_j \bar{x}_{ij} \leq \theta \bar{x}_{io} && \forall i \\
&\quad \text{prob}(\bar{x}_{ij} \in I_{ij}) = K\% && \forall i, \forall j \\
&\quad \sum_{j=1}^{n} \lambda_j \bar{y}_{rj} \geq \bar{y}_{ro} && \forall r \qquad (4) \\
&\quad \text{prob}(\bar{y}_{rj} \in I_{rj}) = K\% && \forall r, \forall j \\
&\quad \lambda_j \geq 0 && \forall j
\end{aligned}
$$

In the above model $I_{ij} = [x_{ij}^{l}, x_{ij}^{u}]$ and $l_{rj} = [y_{rj}^{l}, y_{rj}^{u}]$ are respectively the k% confidence intervals of the inputs and outputs ofDMU_j. By replacing $I_{ij} = [x_{ij}^{l}, x_{ij}^{u}]$ and $l_{rj} = [y_{rj}^{l}, y_{rj}^{u}]$, the model has following format:

$$
\begin{aligned}
&\text{Min } \theta_o \\
&S.t. \quad \sum_{j=1}^{n} \lambda_j \bar{x}_{ij} \leq \theta \bar{x}_{io} && \forall i \\
&\quad \text{prob}(\bar{x}_{ij} \in [x_{ij}^{l}\ x_{ij}^{u}]) = K\% && \forall i, \ \forall j \\
&\quad \sum_{j=1}^{n} \lambda_j \bar{y}_{rj} \geq \bar{y}_{ro} && \forall r \qquad (5) \\
&\quad \text{prob}(\bar{y}_{rj} \in [y_{rj}^{l}\ y_{rj}^{u}]) = K\% && \forall r, \ \forall j \\
&\quad \lambda_j \geq 0 && \forall j
\end{aligned}
$$

As model (5) shows the inputs and output vectors are inserted in an interval with possibility of k%. Model (5) is introduced for interval data. The argument here is emphasizing on historical data. That is, data set for N-th period are set in the confidence interval using the historical data of N-1 periods. Toward to this end, an algorithm is proposed with three steps as follows:

3.1. Confidence interval algorithm:

Step 1. Determine confidence interval α for N-th period by ITSM software with N-1 inputs and outputs of the previous period for n decision-making units. For reliable outcome, α = k% was considered.
Step 2. Impose management opinion on inputs and outputs concerning experience, expertise, familiarity with the workplace, performance, and history of units.
Step 3. Share confidence interval for inputs and outputs of step 1 with a confidence interval for step 2.
In step1, the purpose is forecasting the future based on available data with the least possible error. Therefore, we remove data flow, stabilize variance by using available transformations, and identify the initial model and finally forecast α = k% confidence interval for inputs and outputs. The above algorithm

provides a confidence interval for uncertain input or output with confident of $\alpha = k\%$. After this step, with reference to these intervals, the next step comes to efficiency evaluation. In this step, an interval efficiency is minded.

3.2. Proposed model

Confidence interval discards method

In this method, first the confidence interval of inputs and outputs was divided into k equal parts and then in each of the sub-intervals the inputs and outputs were considered as a convex combination of their lower and upper borders. For each of the decision-making units, k was the efficiency number. The mean of this k number was obtained as the efficiency of the decision-making units. In this way, the proposed model is presented as follows:

$$\begin{aligned}
\text{Max} \quad & \sum_{r=1}^{s} u_r \hat{y}_{ro} \\
S.t. \quad & \sum_{i=1}^{m} v_i \hat{x}_{io} = 1, \\
& \sum_{r=1}^{s} u_r \hat{y}_{rj} - \sum_{i=1}^{m} v_i \hat{x}_{ij} \le 0, && \forall j \\
& \hat{x}_{ij} = x_{ij}^{l} + \left(x_{ij}^{u} - x_{ij}^{l}\right)t, && \forall i, \forall j \\
& y_{rj} = y_{rj}^{l} + \left(y_{rj}^{u} - y_{rj}^{l}\right)t, && \forall r, \forall j \\
& v_i \ge 0, && \forall i, \\
& u_r \ge 0, && \forall r.
\end{aligned} \tag{6}$$

In the above model, $\hat{x}_{ij}$ and و$\hat{y}_{rj}$ are non-negative values and are obtained from the relation $\hat{x}_{ij} = x_{ij}^{l} + \left(x_{ij}^{u} - x_{ij}^{l}\right)t$. Where t is a parameter with a certain value between zero and one. Then the proposed algorithm to solve the proposed model (5) is presented as follows:

Step1: Divide $\alpha = k\%$ of the inputs and outputs into equal parts by k + 1

Step2: Consider$\left\{0, \frac{1}{k}, \frac{2}{k}, \dots, 1\right\}$ and for each value of t, $\hat{x}_{ij}$ and $\hat{y}_{rj}$ is calculated, then solve the corresponding values of model (6). That is, model (6) is calculated for (k + 1) different values of $\hat{x}_{ij}$ and $\hat{y}_{rj}$. And the efficiency of the unit under evaluation is calculated. In this way (k + 1) the efficiency number is obtained.

Step 3: The average (k + 1) efficiency number obtained by each of the decision-making units from step 2 should be considered as the final efficiency number.

The methodology for determining the validity of the proposed model is as follows:

1. We forecast the 95% confidence interval of the 50th inputs and outputs with the help of Time series and through ITSM software, and using actual inputs and outputs of 49 previous periods.
2. We will consider the manager's suggestion for inputs and outputs based on the records of the previous 49 periods for the 50th period, and we will share it with the interval which has been obtained in step one.
3. The efficiency of period 50 is forecasted using proposed model (5). Table 5 shows the shared input and output intervals of step 2 and the efficiency of the 50th period, and also the CCR efficiency of the actual 50th period.
4. We use the rank sum test to validate the proposed model. In fact, we compare the calculated efficiency with the actual efficiency of the 50th period.

5. We apply steps 1, 2 and 3 for Monte Carlo solution methods and interval model. Table 6 shows the upper and lower bounds, and also the average efficiency of the two methods.
6. Comparing above three methods, the best solution method is the proposed model. Table 7 shows the comparison of results using the rank sum test.
After determining the validity of the model, we use steps 1 and 2 to forecast the efficiency of the period 51 by using the discretization of the confidence interval, the results of which are presented in Table 8.
The proposed methodology looks simple to follow. As the algorithms admits the confidence interval are generated. But as the algorithm presents the future data set can be inserted in an interval with k% confidence. This confident allows the manager to focus on their production procedure. That is, the decision maker is satisfied, since their perspective lead to these results. Based on these reliable results, the efficiency for N-th period is calculated. This outcome is trustable and the production procedure can be justified with the k% confidence. Furthermore, a validation test is done until the calculated efficiency can be compared.

4. NUMERICAL EXAMPLE

In order to shed a light on the applicability of proposed algorithm, one of the main industries in Iran is selected. The automobile industry is one of the important industries in the country also identifying some factors can affect the efficiency or inefficiency of this industry. Some of these factors are listed as conditions of business, certain strategy, paying attention to research and development, required liquidity and budget and having a realistic vision according to capabilities. Moreover, if there is no control and realistic vision about the future performance for this industry, the industry will face various challenges. Therefore, forecasting future performance of this industry can play an important role in preventing loss and decreasing risk of financial and human resources. Thus, management can make a long-term plan for its performance and design plans for improving the management of costs and increasing efficiency. As a practical example of the proposed model, 20 automobile industries are selected. The inputs and outputs are collected seasonally and related to 50 time periods. The inputs are number of personnel (x_1) and number of equipment (x_2). The outputs are included Value Added(y_1) and average employee productivity(y_2).
As proposed confidence interval algorithm suggests, a three steps algorithm provides a confidence interval as follows:
Step1: Inputs and outputs of previous 49 periods of these 20 industrial units along with α = 95% are used to provide the inputs and outputs of 50^{th} period. The obtained confidence interval for inputs and outputs are inserted employing ITSM software. These steps are repeated for all inputs and outputs for 20 industrial units. For example, assume the first input of DMU_5. Initial data are shown as follows in Table (1) and Figure (5).

Table 1. The first input of DMU_5 for 49 periods of 20 decision-making units

step1	step2	step3	step4	step5	step6	step7	step8	step9	step10
88.06	87.63	88.07	90.29	81.35	90.02	89.72	83.55	86.15	90.69
step11	step12	step13	step14	step15	step16	step17	step18	step19	step20
86.47	83.22	80.7	87.57	80.92	91.45	83.16	89.57	85.61	91.2
step21	step22	step23	step24	step25	step26	step27	step28	step29	step30
89.12	82.94	83.96	89.74	90.67	87.23	85.23	90.5	86.25	89.19
step31	step32	step33	step34	step35	step36	step37	step38	step39	step40

86.9	86.62	88.82	90.8	87.27	82.26	85.19	87.13	88.97	86.16

step 41	step 42	step 43	step 44	step 45	step 46	step 47	step 48	step 49	Step 50
87.7	89.74	84.22	80.44	83.2	87.29	91.03	87.27	88.13	80.06

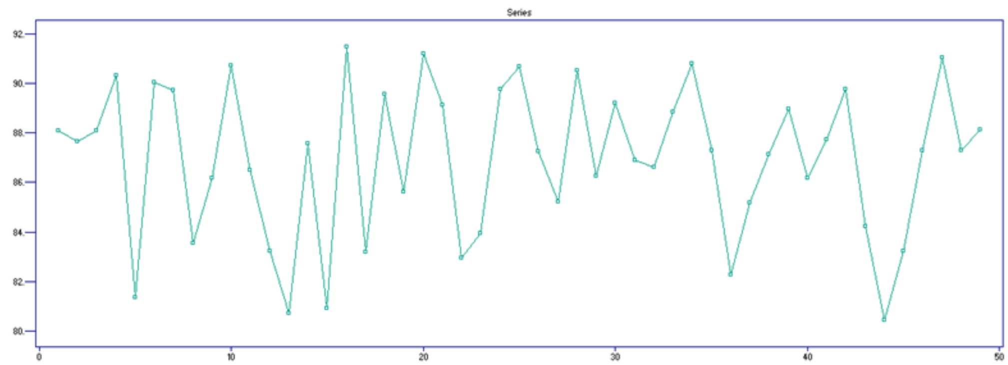

Figure 5. Graphical representation of table1

Reduction the scattering and differentiation of the data for unit#5, Figure (6) is derived.

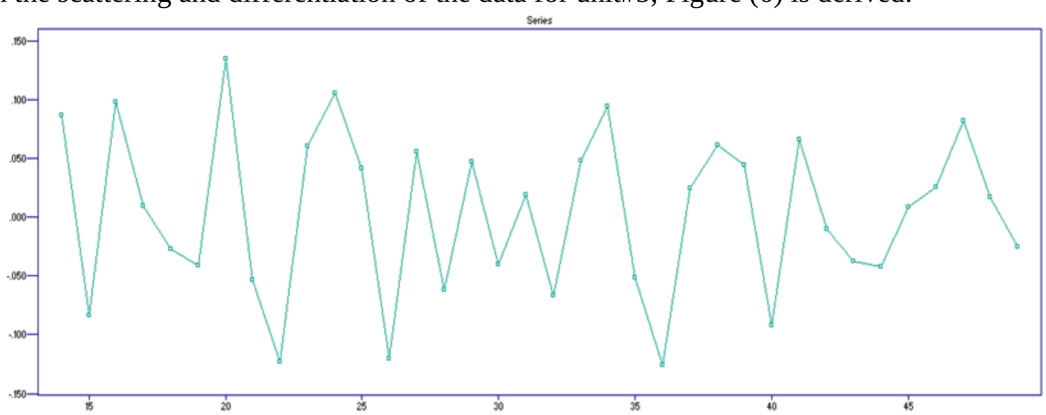

Figure 6. Decreasing dispersion and differentiating in first input of DMU_5

Subsequently, the used software make data fit with the appropriate model to forecast inputs and outputs for 50^{th} period. After determining the appropriate model based on available data, inputs and outputs are forecasted with desirable possibility at a 95% confidence interval. Figure (7) shows the results.

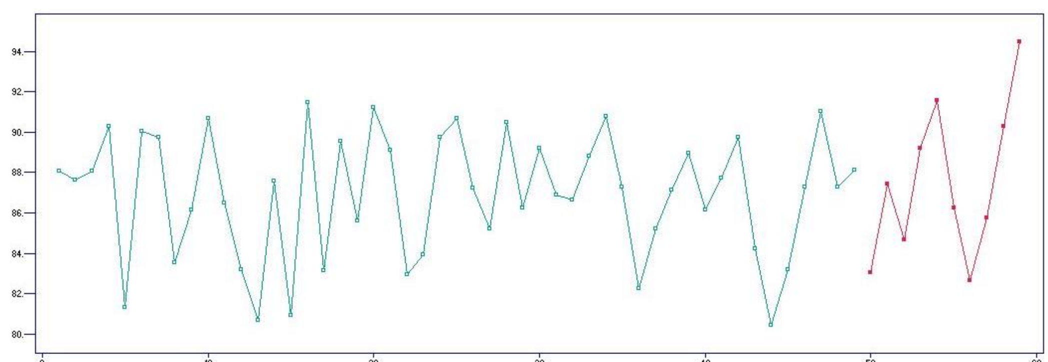

Figure 7. Forecasting 50^{th} to59 th periods for the first input of DMU_5

Table (2) shows the related resulted of forecasting 95% confidence interval for inputs and outputs for 50th to 59th periods for the first input of DMU_5.

Table 2. Results of forecasting 50th to 59th periods for the first input of DMU_5

		Approximate 95 Percent Prediction Bounds	
step	prediction	Lower	Upper
50	83.04	76.42	90.23
51	87.45	79.85	95.78
52	84.66	77.28	92.74
53	89.20	80.91	98.33
54	91.58	82.45	101.72
55	86.24	77.09	96.47
56	82.64	73.39	93.06
57	85.77	75.68	97.19
58	90.28	79.19	102.93
59	94.46	82.39	108.92

The first step of confidence algorithm process are done for all units and results are presented as in Table (3).

Table3. Forecasting the confidence interval of 95% of inputs and outputs

DMU	I_1	I_2	O_1	O_2
DMU_1	[76.4 103.55]	[60.28 80.70]	[37.48 51.00]	[47.62 76.63]
DMU_2	[78.11 98.89]	[57.68 78.23]	[40.59 55.28]	[49.28 80.59]
DMU_3	[72.17 88.84]	[71.13 92.70]	[40.59 55.38]	[45.56 64.63]
DMU_4	[80.52 103.85]	[71.57 99.79]	[38.82 52.35]	[37.00 55.78]
DMU_5	[76.42 90.23]	[60.90 86.53]	[34.68 46.52]	[45.93 65.63]
DMU_6	[72.88 95.8]	[61.95 81.05]	[40.95 62.11]	[47.64 69.58]
DMU_7	[84.24 98.68]	[70.34 92.48]	[43.38 55.62]	[44.16 68.91]
DMU_8	[82.21 98.06]	[63.41 95.10]	[44.90 61.53]	[42.95 65.59]
DMU_9	[82.15 101.91]	[71.69 105.61]	[37.74 53.19]	[35.64 57.92]
DMU_{10}	[82.73 101.48]	[83.69 112.43]	[39.59 51.91]	[40.85 63.69]
DMU_{11}	[83.19 99.42]	[71.95 102.52]	[37.2 50.98]	[31.42 55.29]
DMU_{12}	[75.46 93.44]	[69.37 95.91]	[39.71 63.52]	[38.98 57.01]
DMU_{13}	[72.18 90.48]	[71.30 90.09]	[43.02 59.88]	[50.87 81.03]
DMU_{14}	[76.50 95.92]	[57.68 81.76]	[27.91 48.83]	[44.53 63.99]
DMU_{15}	[79.88 100.39]	[61.46 84.79]	[38.66 54.05]	[45.33 68.69]
DMU_{16}	[75.94 93.02]	[57.74 79.56]	[40.55 53.98]	[50.57 72.96]
DMU_{17}	[74.65 93.15]	[73.73 97.53]	[32.83 44.94]	[46.83 73.92]
DMU_{18}	[74.36 88.89]	[66.86 91.21]	[39.36 52.45]	[47.98 80.81]
DMU_{19}	[83.12 104.39]	[64.18 88.25]	[40.83 61.70]	[40.38 62.24]
DMU_{20}	[74.93 91.92]	[61.84 83.21]	[40.31 58.21]	[43.19 68.37]

Coming to step 2 of confidence interval algorithm, imposing management opinion on inputs and outputs, the results are shown in Table(4).

Table 4. Inputs and outputs by applying management feedback

DMU	I_1	I_2	O_1	O_2
DMU_1	[72.35 88.42]	[78.25 86.48]	[45.66 49.47]	[48.99 62.35]
DMU_2	[78.33 91.95]	[75.91 85.59]	[38.38 42.4]	[43.85 56.95]
DMU_3	[77.55 92.87]	[77.88 89.60]	[40.42 45.58]	[44.97 59.61]
DMU_4	[78.57 92.23]	[74.6 84.12]	[42.24 50.58]	[42.71 57.77]
DMU_5	[79.32 96.94]	[74.44 82.04]	[41.71 44.27]	[45.28 55.34]
DMU_6	[77.52 89.18]	[75.9 87.32]	[41.88 50.16]	[42.79 56.71]
DMU_7	[78.06 86.2]	[65.71 75.59]	[41.91 50.19]	[49.33 60.29]
DMU_8	[83.22 93.84]	[63.45 77.55]	[47.28 52.24]	[49.65 60.67]
DMU_9	[84.03 91.03]	[74.28 80.46]	[38.72 44.54]	[51.02 59.88]
DMU_{10}	[78.06 95.40]	[70.83 83.13]	[41.35 44.79]	[49.77 60.81]
DMU_{11}	[80.65 87.37]	[64.45 75.65]	[44.96 47.74]	[45 54.98]
DMU_{12}	[78.85 87.15]	[69.01 77.81]	[44.29 50.95]	[53.25 62.51]
DMU_{13}	[79.61 93.45]	[74.87 82.75]	[42.81 51.27]	[51.05 62.39]
DMU_{14}	[83.2 95.72]	[76.36 84.38]	[42.41 46.65]	[43 53.62]
DMU_{15}	[73.35 89.65]	[75.37 88.47]	[39.98 44.18]	[48.77 58.41]
DMU_{16}	[79.75 91.75]	[69.19 76.47]	[44.20 46.92]	[53.74 63.08]
DMU_{17}	[78.36 86.60]	[77.9 87.84]	[44.63 51.33]	[42.67 50.09]
DMU_{18}	[80.16 92.22]	[73.99 81.77]	[41.91 51.21]	[44.73 53.57]
DMU_{19}	[82.7 91.40]	[76.4 82.76]	[44.07 48.69]	[47.11 59.95]
DMU_{20}	[78.6 88.62]	[69.3 75.06]	[43.42 50.96]	[48.31 59.03]

In the following step, step3, the confidence interval for inputs and outputs of step 1 are being shared with a confidence interval for step 2. The participated inputs and outputs for 50th period are depicted in table (5).

Table 5. The inputs and outputs confidence intervals and the actual and forecasted efficiency of the 50th period

DMU	I_1	I_2	O_1	O_2	*Estimated Efficiency in Period 50*	*The actul Efficiey of the 50th*
DMU_1	[76.48 88.42]	[78.25 80.70]	[45.66 49.47]	[48.99 62.35]	1	0.88
DMU_2	[78.33 91.95]	[75.91 78.23]	[40.59 42.4]	[49.28 56.95]	0.91	0.97
DMU_3	[77.55 88.84]	[77.88 89.60]	[40.59 45.58]	[45.56 59.61]	0.92	0.97
DMU_4	[80.52 92.23]	[74.6 84.12]	[42.24 50.58]	[42.71 55.78]	0.92	0.88
DMU_5	[79.32 90.23]	[74.44 82.04]	[41.71 44.27]	[45.93 55.34]	0.88	1
DMU_6	[77.52 89.18]	[75.9 81.05]	[41.88 50.16]	[47.64 56.71]	0.95	0.94
DMU_7	[78.06 86.26]	[70.34 75.59]	[43.38 50.19]	[49.33 60.29]	0.99	0.92
DMU_8	[83.22 93.84]	[63.45 77.55]	[47.28 52.24]	[49.65 60.67]	1	1

DMU_9	[84.03 91.03]	[74.28 80.46]	[33.72 44.54]	[51.02 57.92]	0.91	1
DMU_{10}	[82.73 95.40]	[70.83 83.13]	[41.35 44.79]	[49.77 60.81]	0.91	0.81
DMU_{11}	[83.19 87.37]	[71.95 75.65]	[44.96 47.74]	[45 54.98]	0.94	0.98
DMU_{12}	[78.85 87.15]	[69.37 77.81]	[44.29 50.95]	[53.25 57.01]	1	0.94
DMU_{13}	[79.61 90.48]	[74.87 82.75]	[43.02 51.27]	[51.05 62.39]	0.98	0.99
DMU_{14}	[83.2 95.72]	[76.36 81.76]	[42.41 46.65]	[44.53 53.62]	0.86	0.88
DMU_{15}	[79.88 89.65]	[75.37 84.79]	[39.98 44.18]	[48.77 58.41]	0.92	1
DMU_{16}	[79.75 91.75]	[69.19 76.47]	[44.20 46.92]	[53.74 63.08]	1	0.92
DMU_{17}	[78.36 86.60]	[77.9 87.84]	[44.63 44.94]	[46.83 50.09]	0.93	0.96
DMU_{18}	[80.16 88.89]	[73.99 81.77]	[41.91 51.21]	[47.98 53.57]	0.95	0.93
DMU_{19}	[83.12 91.40]	[76.4 82.76]	[44.07 48.69]	[47.11 59.95]	0.92	1
DMU_{20}	[78.6 88.62]	[69.3 75.06]	[43.42 50.96]	[48.31 59.03]	0.98	0.94

By implementing step 1, step 2 and sep3 of the proposed algorithm, the results are presented in table (6).

Table 6. Results of the efficiency of the Monte Carlo and interval methods for forecasting the 50th period

DMU_j	*Efficiency by Monte Carlo method*				*Interval efficiency*		
	θ^l	θ^u	θ^{ave}	STD	θ^l	θ^u	θ^{ave}
DMU_1	0.75	0.96	0.85	0.04	0.79	1	0.89
DMU_2	0.85	1	0.96	0.03	0.70	1	0.85
DMU_3	0.74	0.99	0.85	0.05	0.70	1	0.85
DMU_4	0.70	0.86	0.78	0.03	0.70	1	0.85
DMU_5	0.68	0.94	0.81	0.05	0.71	1	0.85
DMU_6	0.86	1	0.96	0.03	0.72	1	0.86
DMU_7	0.80	1	0.95	0.04	0.77	1	0.88
DMU_8	0.86	1	0.98	0.03	0.82	1	0.91
DMU_9	0.76	1	0.90	0.05	0.70	1	0.85
DMU_{10}	0.68	0.87	0.79	0.03	0.68	1	0.84
DMU_{11}	0.69	0.90	0.81	0.04	0.79	1	0.89
DMU_{12}	0.84	1	0.96	0.03	0.79	1	0.89
DMU_{13}	0.93	1	0.99	0.01	0.73	1	0.86
DMU_{14}	0.65	0.81	0.72	0.03	0.68	1	0.84
DMU_{15}	0.82	1	0.93	0.05	0.69	1	0.84
DMU_{16}	0.79	1	0.90	0.04	0.77	1	0.88
DMU_{17}	0.72	0.89	0.80	0.03	0.79	1	0.89
DMU_{18}	1	1	1	0.00	0.72	1	0.86
DMU_{19}	0.88	1	0.96	0.03	0.74	1	0.87
DMU_{20}	0.78	0.97	0.88	0.04	0.76	1	0.88

Using the rank sum test, the value of T is obtained from the following equation.

$$T = \frac{S - {m(m+n+1)}/{2}}{\sqrt{{mn(m+n+1)}/{12}}}$$

Where m is the number of data in the first set and n is the number of data in the second set and also the statistic s approximately follows the normal distribution with mean value of $\frac{m(m+n+1)}{2}$ and the variance of $\frac{mn(m+n+1)}{12}$.

Table 7. Comparison of the results of the rank sum test

Method	Results
Discretening the confidence interval	$-1.96 \leq T = -0.\mathbf{95} \leq 1.96$
Monte-Carlo	$T = -2.\mathbf{05} \ngeq -1.96$
Interval	$T = -4.\mathbf{36} \ngeq -1.96$

The results of Table 7 indicate that there is no significant difference in the assumption of the rejection of the actual efficiency inconsistency and the efficiency which has been forecasted by the confidence interval method. That is, the forecasted efficiency numbers have computational desirability. Therefore, the confidence interval method is the best way to forecast efficiency for future periods.

Now, considering the determination of the validity of the proposed model, we used the data of the previous 50 periods to forecast the efficiency of period 51, the results of which are presented in Table 8.

Table 8. Forecasted results of period 51 using the confidence interval discard method

DMU_j	*Estimated Efficiency in Period 51*	*The actual Efficiency of the 50th*
DMU_1	0.90	0.88
DMU_2	0.97	0.97
DMU_3	0.98	0.97
DMU_4	0.87	0.88
DMU_5	1	1
DMU_6	0.92	0.94
DMU_7	0.85	0.92
DMU_8	0.98	1
DMU_9	1	1
DMU_{10}	0.83	0.81
DMU_{11}	0.98	0.98
DMU_{12}	0.95	0.94
DMU_{13}	0.98	0.99
DMU_{14}	0.90	0.88
DMU_{15}	1	1
DMU_{16}	0.94	0.92
DMU_{17}	0.99	0.96
DMU_{18}	0.94	0.93
DMU_{19}	1	1
DMU_{20}	0.92	0.94

The data in Table 8 includes the actual efficiency of the 50th period and the forecasted efficiency of the 51st period. These results indicate that with the probability of 95%, the fifth, ninth, fifteenth and nineteenths industrial units will remain efficient in the next period, and units like the second and eleventh will also maintain their previous efficiency. Among the industrial units, the seventh industrial unit in the 50th period has $\theta_7^* = 0.92$ and in the period 51 has$\theta_7^* = 0.85$, which will reduce the efficiency of this unit by about 8%. Therefore, in order to investigate the causes and factors of reducing the efficiency, the results of the forecast are announced to the manager in order to take the necessary measures.

5. CONCLUSION

Data envelopment analysis models are used to evaluate the efficiency of decision making units with known inputs and outputs. One of the major drawbacks of standard DEA models is being retrospective and thus these models cannot forecast. Obviously, the management of each decision-making unit is interested in forecasting the efficiency interval in the upcoming periods, so that it regulates its activities in terms of resource consumption and output generation, but there are also situations where the data is definitely not in the given interval.

In this paper, we forecasted inputs and outputs by creating confidence intervals of 95% using time series. Then we used by 95% confidence intervals in a proposed model as a generalization of IDEA. The proposed model as a forecast model was used to estimate the efficiency and efficiency confidence interval of 20 decision-making units of the industry, with available information for their 50assessment period. We use confidence interval discard, Monte Carlo simulation, and interval methods to solve this model.

To validate these solution methods, we used the data of the previous 49 periods to forecast the efficiency of the 50th period and compared it with the actual efficiency of the 50th period by means of rank sum test. The results show that there is not a significant difference between the actual efficiency of the 50th period and the forecasted results by confidence interval discard. Finally, we use the proposed method to forecast the efficiency of the 51st period.

6. REFERENCES

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science*, *30*(9), 1078-1092.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research*, *2*(6), 429-444.

Cooper, W. W., Park, K. S., & Yu, G. (1999). IDEA and AR-IDEA: Models for dealing with imprecise data in DEA. *Management science*, *45*(4), 597-607.

Cooper, W. W., Park, K. S., & Yu, G. (2001). An illustrative application of IDEA (imprecise data envelopment analysis) to a Korean mobile telecommunication company. *Operations Research*, *49*(6), 807-820.

Wang, Y. M., Greatbanks, R., & Yang, J. B. (2005). Interval efficiency assessment using data envelopment analysis. *Fuzzy sets and Systems*, *153*(3), 347-370.

Jahanshahloo, G. R., Lotfi, F. H., Malkhalifeh, M. R., & Namin, M. A. (2009). A generalized model for data envelopment analysis with interval data. *Applied Mathematical Modelling*, *33*(7), 3237-3244.

Fildes, R., Wei, Y., & Ismail, S. (2011). Evaluating the forecasting performance of econometric models of air passenger traffic flows using multiple error measures. *International Journal of Forecasting*, *27*(3), 902-

Xu, B., & Ouenniche, J. (2012). A data envelopment analysis-based framework for the relative performance evaluation of competing crude oil prices' volatility forecasting models. *Energy Economics*, *34*(2), 576-583.

Lim, D. J., Anderson, T. R., & Inman, O. L. (2014). Choosing effective dates from multiple optima in technology forecasting using data envelopment analysis (TFDEA). *Technological Forecasting and Social Change*, *88*, 91-97.

Lim, D. J., Anderson, T. R., & Shott, T. (2015). Technological forecasting of supercomputer development: The march to exascale computing. *Omega*, *51*, 128-135.

Emrouznejad, A., Rostami-Tabar, B., & Petridis, K. (2016). A novel ranking procedure for forecasting approaches using Data Envelopment Analysis. *Technological Forecasting and Social Change*, *111*, 235-243.

Angiz, M. Z., Tajaddini, A., Mustafa, A., & Kamali, M. J. (2012). Ranking alternatives in a preferential voting system using fuzzy concepts and data envelopment analysis. Computers & Industrial Engineering, 63(4), 784-790.

Hatami-Marbini, A., Emrouznejad, A., & Tavana, M. (2011). A taxonomy and review of the fuzzy data envelopment analysis literature: two decades in the making. European journal of operational research, 214(3), 457-472.

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science*, *30*(9), 1078-1092.

Shabanpour, H., Yousefi, S., & Saen, R. F. (2017). Forecasting efficiency of green suppliers by dynamic data envelopment analysis and artificial neural networks. Journal of Cleaner Production, 142, 1098-1107.

Peykani, P., Mohammadi, E., Rostamy-Malkhalifeh, M., & Hosseinzadeh Lotfi, F. (2019). Fuzzy data envelopment analysis approach for ranking of stocks with an application to Tehran stock exchange. Advances in Mathematical Finance and Applications, 4(1), 31-43.

Tavassoli, M., & Saen, R. F. (2019). Predicting group membership of sustainable suppliers via data envelopment analysis and discriminant analysis. Sustainable Production and Consumption, 18, 41-52.